AD NO. ___________________________
DTC PROJECT NO. 8-CO-160-UXO-021
REPORT NO. ATC-10523

STANDARDIZED
UXO TECHNOLOGY DEMONSTRATION SITE

BLIND GRID SCORING RECORD NO. 926

SITE LOCATION:
U.S. ARMY YUMA PROVING GROUND

DEMONSTRATOR:
U.S. GEOLOGICAL SURVEY
DENVER FEDERAL CENTER
BLDG. 20, MS-964
DENVER, CO 80225-0046

TECHNOLOGY TYPE/PLATFORM:
ALLTEM/TOWED

PREPARED BY:
U.S. ARMY ABERDEEN TEST CENTER
ABERDEEN PROVING GROUND, MD 21005-5059

MARCH 2011

Prepared for:
SERDP/ESTCP
MUNITIONS MANAGEMENT
ARLINGTON, VA 22203

U.S. ARMY DEVELOPMENTAL TEST COMMAND
ABERDEEN PROVING GROUND, MD 21005-5055

DISTRIBUTION UNLIMITED, MARCH 2011.

DISPOSITION INSTRUCTIONS

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
March 2011	Final	17 through 20, 25, and 27 February 2009

4. TITLE AND SUBTITLE	
STANDARDIZED UXO TECHNOLOGY DEMONSTRATION SITE BLIND GRID SCORING RECORD NO. 926	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) McClung, J. Stephen	**5d. PROJECT NUMBER** 8-CO-160-UXO-021
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Commander U.S. Army Aberdeen Test Center ATTN: TEDT-AT-SLE Aberdeen Proving Ground, MD 21005-5059	ATC-10523

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
Commander U.S. Army Environmental Command ATTN: IMAE-RTA Aberdeen Proving Ground, MD 21010-5401	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)** Same as item 8

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution unlimited.

13. SUPPLEMENTARY NOTES
None

14. ABSTRACT
This scoring record documents the efforts of USGS to detect and discriminate inert unexploded ordnance (UXO) utilizing the YPG Standardized UXO Technology Demonstration Site blind grid. This scoring record was coordinated by J. Stephen McClung and the Standardized UXO Technology Demonstration Site Scoring Committee. Organizations on the committee include the U.S. Army Corps of Engineers, the Environmental Security Technology Certification Program, the Strategic Environmental Research and Development Program, the Institute for Defense Analysis, the U.S. Army Environmental Command, and the U.S. Army Aberdeen Test Center.

15. SUBJECT TERMS

Blind grid, ALLTEM/towed

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
Unclassified	Unclassified	Unclassified	SAR		

ACKNOWLEDGMENTS

Authors:

William Burch
Leonardo Lombardo
Homeland Defense and Sustainment Division (HDSD)
U.S. Army Aberdeen Test Center
Aberdeen Proving Ground

Rick Fling
Aberdeen Test Support Services (ATSS)
Sverdrup Technology, Inc.
Aberdeen Proving Ground (APG)

Contributors:

J. Stephen McClung
Homeland Defense and Sustainment Division (HDSD)
U.S. Army Aberdeen Test Center
Aberdeen Proving Ground

Christina McClung
Survivability and Lethality (SL) Directorate
U.S. Army Aberdeen Test Center (ATC)
Aberdeen Proving Ground

TABLE OF CONTENTS

SECTION 1. GENERAL INFORMATION

1.1 BACKGROUND

Technologies under development for the detection and discrimination of unexploded ordnance (UXO) require testing so that their performance can be characterized. To that end, Standardized Test Sites have been developed at Aberdeen Proving Ground (APG), Maryland, and U.S. Army Yuma Proving Ground (YPG), Arizona. These test sites provide a diversity of geology, climate, terrain, and weather as well as diversity in ordnance and clutter. Testing at these sites is independently administered and analyzed by the Government for the purposes of characterizing technologies, tracking performance with system development, comparing performance of different systems, and comparing performance in different environments.

The Standardized UXO Technology Demonstration Site Program is a multiagency program spearheaded by the U.S. Army Environmental Command (USAEC). The U.S. Army Aberdeen Test Center (ATC) and the U.S. Army Corps of Engineers Engineering Research and Development Center (ERDC) provide programmatic support. The program is being funded and supported by the Environmental Security Technology Certification Program (ESTCP), the Strategic Environmental Research and Development Program (SERDP), and the Army Environmental Quality Technology Program (EQT).

1.2 SCORING OBJECTIVES

The objective in the Standardized UXO Technology Demonstration Site Program is to evaluate the detection and discrimination capabilities of a given technology under various field and soil conditions. Inert munitions and clutter items are positioned in various orientations and depths in the ground.

The evaluation objectives are as follows:

a. To determine detection and discrimination effectiveness under realistic scenarios that vary targets, geology, clutter, topography, and vegetation.

b. To determine cost, time, and manpower requirements to operate the technology.

c. To determine the demonstrator’s ability to analyze survey data in a timely manner and provide prioritized “Target Lists” with associated confidence levels.

d. To provide independent site management to enable the collection of high quality, ground-truth, geo-referenced data for post-demonstration analysis.

1.2.1 <u>Scoring Methodology</u>

a. The scoring of the demonstrator’s performance is conducted in two stages. These two stages are termed the RESPONSE STAGE and DISCRIMINATION STAGE. For both stages, the probability of detection (P_d) and the false alarms are reported as receiver-operating characteristic (ROC) curves. False alarms are divided into those anomalies that correspond to emplaced clutter items, measuring the probability of false positive (P_{fp}), and those that do not correspond to any known item, termed background alarms.

b. The RESPONSE STAGE scoring evaluates the ability of the system to detect emplaced targets without regard to ability to discriminate ordnance from other anomalies. For the blind grid RESPONSE STAGE, the demonstrator provides the scoring committee with a target response from each and every grid square along with a noise level below which target responses are deemed insufficient to warrant further investigation. This list is generated with minimal processing and, since a value is provided for every grid square, will include signals both above and below the system noise level.

c. The DISCRIMINATION STAGE evaluates the demonstrator's ability to correctly identify ordnance as such and to reject clutter. For the blind grid DISCRIMINATION STAGE, the demonstrator provides the scoring committee with the output of the algorithms applied in the discrimination-stage processing for each grid square. The values in this list are prioritized based on the demonstrator's determination that a grid square is likely to contain ordnance. Thus, higher output values are indicative of higher confidence that an ordnance item is present at the specified location. For digital signal processing, priority ranking is based on algorithm output. For other discrimination approaches, priority ranking is based on human (subjective) judgment. The demonstrator also specifies the threshold in the prioritized ranking that provides optimum performance (i.e., that is expected to retain all detected ordnance and rejects the maximum amount of clutter).

d. The demonstrator is also scored on EFFICIENCY and REJECTION RATIO, which measures the effectiveness of the discrimination stage processing. The goal of discrimination is to retain the greatest number of ordnance detections from the anomaly list, while rejecting the maximum number of anomalies arising from non-ordnance items. EFFICIENCY measures the fraction of detected ordnance retained after discrimination, while the REJECTION RATIO measures the fraction of false alarms rejected. Both measures are defined relative to performance at the demonstrator-supplied level below which all responses are considered noise, i.e., the maximum ordnance detectable by the sensor and its accompanying false positive rate or background alarm rate.

e. All scoring factors are generated utilizing the Standardized UXO Probability and Plot Program, version 3.1.1.

1.2.2 Scoring Factors

Factors to be measured and evaluated as part of this demonstration include:

a. Response Stage ROC curves:

(1) Probability of Detection (P_d^{res}).

(2) Probability of False Positive (P_{fp}^{res}).

(3) Background Alarm Rate (BAR^{res}) or Probability of Background Alarm (P_{BA}^{res}).

b. Discrimination Stage ROC curves:

(1) Probability of Detection (P_d^{disc}).

(2) Probability of False Positive (P_{fp}^{disc}).

(3) Background Alarm Rate (BAR^{disc}) or Probability of Background Alarm (P_{BA}^{disc}).

c. Metrics:

(1) Efficiency (E).

(2) False Positive Rejection Rate (R_{fp}).

(3) Background Alarm Rejection Rate (R_{BA}).

d. Other:

(1) Probability of Detection by Size and Depth.

(2) Classification by type (i.e., 20-, 40-, 105-mm, etc.).

(3) Location accuracy.

(4) Equipment setup, calibration time, and corresponding man-hour requirements.

(5) Survey time and corresponding man-hour requirements.

(6) Reacquisition/resurvey time and man-hour requirements (if any).

(7) Downtime due to system malfunctions and maintenance requirements.

1.3 STANDARD AND NONSTANDARD INERT ORDNANCE TARGETS

The standard and nonstandard ordnance items emplaced in the test areas are listed in Table 1. Standardized targets are members of a set of specific ordnance items that have identical properties to all other items in the set (caliber, configuration, size, weight, aspect ratio, material, filler, magnetic remanence, and nomenclature). Nonstandard targets are inert ordnance items having properties that differ from those in the set of standardized targets.

TABLE 1. INERT ORDNANCE TARGETS

Standard Type	Nonstandard (NS)
20-mm Projectile M55	20-mm Projectile M55
	20-mm Projectile M97
40-mm Grenades M385	40-mm Grenades M385
40-mm Projectile MKII Bodies	40-mm Projectile M813
BDU-28 Submunition	
BLU-26 Submunition	
M42 Submunition	
57-mm Projectile APC M86	
60-mm Mortar M49A3	60-mm Mortar (JPG)
	60-mm Mortar M49
2.75-inch Rocket M230	2.75-inch Rocket M230
	2.75-inch Rocket XM229
MK 118 ROCKEYE	
81-mm Mortar M374	81-mm Mortar (JPG)
	81-mm Mortar M374
105-mm HEAT Rounds M456	
105-mm Projectile M60	105-mm Projectile M60
155-mm Projectile M483A1	155-mm Projectile M483A
	500-lb Bomb
	M75 Submunition

HEAT = high-explosive antitank.
JPG = Jefferson Proving Ground.

SECTION 2. DEMONSTRATION

2.1 DEMONSTRATOR INFORMATION

2.1.1 Demonstrator Point of Contact (POC) and Address

POC: Ted Asch
(303) 236-2489

Address: U.S. Geological Survey
Denver Federal Center
Bldg. 20, MS-964
Denver, CO 80225-0046

2.1.2 System Description (provided by demonstrator)

The ALLTEM is an 'on-time' time-domain electromagnetic induction system that consists of exciting and detecting 3-component fields using multiple Tx and Rx coils. The triangle current excitation waveform (pulse rate 90 Hz) provides immediate visual separation between ferrous and non-ferrous metal objects. The ALLTEM records data to late times which helps suppress the geologic response relative to the UXO response. The system is pulled by a small Kubota tractor with a small 2 kW generator at the front (see photo below). The ALLTEM sensor 1-meter cube sits in a cart that has a minimum height above the ground of about 6 inches and can be raised up an additional 6 inches to traverse over surface obstacles. A LEICA GPS1200 RTK system provides the sensor location and also input to a USGS-developed survey navigation program. Survey traverses will have 0.5 meter separation with a data density of approximately 15 to 20 cm (traveling at a nominal speed of 0.5 m/sec with a sampling cycle rate for each Tx coil of approximately 300 ms).

Figure 1.   ALLTEM/towed.

2.1.3 Data Processing Description (provided by demonstrator)

Target selection criteria: This section will detail the target selection criteria and the data required to implement the criteria by answering the following questions:

a. *What kind of pre-processing (if any) is applied to the raw data (e.g., filtering, etc)?*ALLTEM preprocessing is a batch process of all binary waveform survey data via a LABVIEW program that performs background subtraction, low-pass and band-width filtering, determination of ferrous/nonferrous/mixed composition, and then exports an ASCII file containing data at 16 time gates along the waveform.

b. *What is the format of the data both pre and post processing of the raw data (e.g., ASCII, binary, etc)?* The original LABVIEW acquisition data consists of binary waveform files with ASCII headers. There is one file per configuration. These are converted in the LABVIEW preprocessing program to an ASCII format that is carried throughout the rest of the processing and analysis.

c. *What algorithm is used for detection (e.g., peaks of signal surpassing threshold, etc)?* In 2008 we have migrated all our processing and analysis software to work within the Geosoft Oasis Montaj platform. Once the data is imported into Oasis, an area that is deemed to be target free is designated. This area forms the threshold basis on which a statistical analysis is performed using the "R Project for Statistical Computing" (http://www.r-project.org/) statistics package.

Wilkes-Shapiro and T-tests characterize the acquired data and then Blakely peakedness tests are performed to designate the locations of the potential targets. This is all done automatically for all 19 ALLTEM receiver configurations.

d. *Why is this algorithm used and not others?* We use the calculated statistics for both picking out targets and as part of the classification analysis at the end of the process.

e. *On what principles is the algorithm based (e.g,. statistical models, heuristic rules, etc.)?* As just mentioned, the algorithm is based on a statistical analysis of the acquired data.

f. *What tunable parameters (if any) are used in the detection process (e.g., threshold on signal amplitude, window length, filter coefficients, etc.)?* Tunable parameters include the background threshold level, the number of standard deviations away from the target threshold used to determine signal levels, the search radius around each selected target (used for merging multiple targets at same location from different receiver polarizations), the areas of what are considered to be statistically 'significant' data for a particular target, and analytic signal calculations for certain receiver polarizations.

g. *What are the final values of all tunable parameters for the detection algorithm?* The final values for the tunable parameters will be determined by the data in the field. The background threshold values will be determined independently for each area surveyed. The search radius will be determined by the largest target detected in each survey area.

Parameter estimation: This section should include the details of which parameters will be extracted from the sensor data for each detected item for characterization. Please answer the following questions:

a. *Which characteristics will be extracted from each detected item and input to the discrimination algorithm (e.g., depth, size, polarizability coefficients, fit quality, etc.)?* Characteristics extracted for each detected item include inferred composition (ferrous/nonferrous/mixed), horizontal location and depth, azimuth, inclination, magnetic polarizability coefficients, and the ratio of polarizability coefficients.

b. *Why have these characteristics been chosen and not others (e.g., empirical evidence of their ability to help discriminate, inclusion in a theoretical tradition, etc.)?* We have determined empirically from previous surveys and by models that these characteristics (composition, polarizabilities, ratios of polarizabilities) have proven effective at discriminating UXO versus clutter versus blank holes.

c. How are these characteristics estimated (e.g., least-mean-squares fit to a dipole model, etc.), include the equations that are used for parameter estimation?

The ALLTEM UXO forward operator approximates the induced field response of a subsurface UXO. This operator describes the induced magnetic fields in the UXO in terms of three orthogonal principal polarizabilities. A set of principal polarizabilities is used to describe the induced magneto-static response. The forward operator A used in the inversion has the form

$$A([\mathbf{r}_{cart},\mathbf{r}_{tx},\mathbf{r}_{rx1},\mathbf{r}_{rx2}\mathbf{P}_{tx},\mathbf{P}_{rx},t],[\mathbf{r}_{s,uxo},\phi_s,\theta_s,m_{s,1},m_{s,2},m_{s,3}]) = y([\mathbf{r}_{cart},\mathbf{r}_{tx},\mathbf{r}_{rx1},\mathbf{r}_{rx2},\mathbf{P}_{tx},\mathbf{P}_{rx},t]) \quad (3)$$

where $\mathbf{r}_{\mathrm{cart}}$ is the location of the center of the ALLTEM cart, $\mathbf{r}_{\mathrm{tx}}$, $\mathbf{r}_{\mathrm{rx1}}$, and $\mathbf{r}_{\mathrm{rx2}}$ are the locations of the transmitting and receiving loops, $\mathbf{P}_{\mathrm{tx}}$ and $\mathbf{P}_{\mathrm{rx}}$ are the polarizations of the transmitter and receiver coils, t is time, and y are the simulated data. The UXO parameter set is listed in the second set of square brackets in the argument list of the forward operator, where $\mathbf{r}_{\mathrm{s,uxo}}$ is the location of the UXO, $m_{\mathrm{s,1}}$ through $m_{\mathrm{s,3}}$ are the magnitude of the three orthogonal induced magneto-static principal polarizabilities, ϕ_s, and θ_s are the azimuth and inclination of the m_1 component. For the induced magneto-static response, the strengths of the three principal polarizability components are specified. The attitude of the first principal polarizability ($\hat{\mathbf{m}}_1$) is described in terms of azimuth and inclination from horizontal, the principal polarizability is horizontal ($\hat{\mathbf{m}}_2 = \hat{\mathbf{m}}_1 \times \hat{\mathbf{z}}$), and the attitude of the third principal polarizability is the cross product of the first two ($\hat{\mathbf{m}}_3 = \hat{\mathbf{m}}_1 \times \hat{\mathbf{m}}_2$).

The ALLTEM UXO forward operator approximates the induced field response of a subsurface UXO. The forward model includes the induced magneto-static response at a fixed instant in time. The magneto-static UXO response is modeled as three orthogonal magnetic dipoles. It is assumed that the target and the ALLTEM cart are in a non-magnetic, non-conducting whole space. The modeled magneto-static induction (T) at a receiver coil $\mathbf{B_s}(\mathbf{r})$ is calculated using

$$\mathbf{B_s}(\mathbf{r}) = \overleftrightarrow{\mathbf{k}} \cdot \left[\frac{3\hat{\mathbf{R}}(\mathbf{m}_s \cdot \hat{\mathbf{R}}) - \mathbf{m}_s}{R^3} \right] \qquad \text{(A1)}$$

where $\overleftrightarrow{\mathbf{k}}$ is a calibration matrix, $\mathbf{r}$ is the location of the receiver, $\mathbf{r}'$ is the UXO location, $\mathbf{R} = \mathbf{r} - \mathbf{r}'$, $R = |\mathbf{R}|$, $\hat{\mathbf{R}} = \mathbf{R}/R$, and $\mathbf{m}_s$ is the static induced dipole moment (A-m^2). The static induced dipole moment is given by

$$\mathbf{m}_s = \begin{bmatrix} \mathbf{m}_{s,1}\hat{\mathbf{m}}_{s,1} & \mathbf{m}_{s,1}\hat{\mathbf{m}}_{s,2} & \mathbf{m}_{s,1}\hat{\mathbf{m}}_{s,3} \\ \mathbf{m}_{s,2}\hat{\mathbf{m}}_{s,1} & \mathbf{m}_{s,2}\hat{\mathbf{m}}_{s,2} & \mathbf{m}_{s,2}\hat{\mathbf{m}}_{s,3} \\ \mathbf{m}_{s,3}\hat{\mathbf{m}}_{s,1} & \mathbf{m}_{s,3}\hat{\mathbf{m}}_{s,2} & \mathbf{m}_{s,3}\hat{\mathbf{m}}_{s,3} \end{bmatrix} \cdot \mathbf{H_p}(\mathbf{r}') \quad \text{(A2)}$$

where $\mathbf{H_p}(\mathbf{r}')$ is the primary magnetic field (A/m), and the matrix is the polarizability tensor (m^3) and the three induced magnetic moments are related by

$$\hat{\mathbf{m}}_2 = \hat{\mathbf{m}}_1 \times \hat{\mathbf{z}} \text{ and } \hat{\mathbf{m}}_3 = \hat{\mathbf{m}}_1 \times \hat{\mathbf{m}}_2. \qquad \text{(A3)}$$

The primary field $\mathbf{H_p}(\mathbf{r}')$ at the UXO location is calculated using the Biot-Savart Law (Jackson, 1999) for the 1-meter-square loop transmitting coils.

d. What tunable parameters (if any) are used in the characterization process? (e.g., thresholds on background noise, etc.)? Tunable parameters include all the parameters derived by the inversion process.

Classification: This section should include the details describing the algorithm and associated data and parameters used for discrimination by answering the following questions:

a. What algorithm is used for discrimination (e.g., multi-layer perception, support vector machine, etc.)? The primary algorithm is an analysis of the polarization coefficients and comparison to coefficients for known items including those from the Calibration grid, the ratios of the polarization coefficients, and the inferred composition from the waveform analysis during preprocessing.

b. Why is this algorithm used and not others? This discrimination analysis process has been used successfully for ALLTEM for UXO items.

c. Which parameters are considered as possible inputs to the algorithm? Polarization coefficients, ratios of polarization coefficients, inferred composition, calculated time constant for target items, signal to noise ratios, size of area of target anomaly.

d. What are the outputs of the algorithm (probabilities, confidence levels)? Multiple probabilities of classification with associated confidence levels are derived for a given target item.

e. How is the threshold set to decide where the munitions/non-munitions line lies in the discrimination process? The threshold used to determine UXO vs clutter is based on the ratio of the polarization constants. For a rod-like item, the two smaller constants should be similar and much smaller than the third, much larger, constant. Clutter typically does not follow this pattern.

Training: This section should include the details of how training data is used to make a decision on the likelihood of the anomaly correspondence to munitions. Please answer the following questions:

a. Which tunable parameters have final values that are optimized over a training set of data and which have values that are set according to geophysical knowledge (i.e., intuition, experience, common sense)? Training data is used to tune estimates of location, depth, polarization constants, time decay constants, and composition analysis. Geophysical knowledge comes in when deciding that a rod-like, sphere-like, or disk-like object is a UXO versus a piece of clutter.

(1) For those tunable parameters with final values set according to geophysical knowledge:

(a) What is the reasoning behind choosing these particular values? These shapes (rod, sphere, disk) seem to be the typical type of ordnance used on training ranges.

(b) Why were the final values not optimized over a training set of data? These are, to a large degree, a priori data at a given site.

(2) For those tunable parameters with final values optimized over the training set data:

(a) What training data is used (e.g., all data, a randomly chosen portion of data, etc.)? All available data is utilized to train the inversion and classification algorithms.

(b) What error metric is minimized during training (e.g., mean squared error, etc.)? Inferred composition analysis and definition of an ordnance by its polarization coefficients and time decay constant.

(c) What learning rule is used during training (e.g., gradient descent, etc.)? Determine best parameters to identify and characterize ordnance versus clutter.

(d) What criterion is used to stop training (e.g., number of iterations exceeds threshold, good generalization over validation set of data, etc.)? Criterion is limit of the number of training items.

(e) Are all tunable parameters optimized at once or in sequence ("in sequence" = parameters 1 is held constant at some common sense values while parameter 2 is optimized, and then parameter 2 is held constant at its optimized value while parameter 1 is optimized)? Tunable parameters are optimized in sequence.

b. What are the final values of all tunable parameters for the characterization process? The final values for the characterization are the correctly classified targets.

2.1.4 Data Submission Format

Data were submitted for scoring in accordance with data submission protocols outlined in the Standardized UXO Technology Demonstration Site Handbook. These submitted data are not included in this report in order to protect ground truth information.

2.1.5 Demonstrator Quality Assurance (QA) and Quality Control (QC) (provided by demonstrator)

Overview of Quality Control (QC): The ALLTEM system has a real-time data display that instantly shows the operator if the transmitting/receiving functions of the system fail. In addition, we plan to find a location with no known targets and repetitively reoccupy that location and record data, including GPS data, to assess and document any drifts that may occur in the instrumentation. Standard operating procedure with all these systems is to occupy a designated clean location at least twice each day: prior to and at the completion of regular data acquisition. This usually takes place in the morning and afternoon, but in case of an extended pause in the middle of the day, an additional reference data set may be acquired. This will also test the accuracy and repeatability of the navigation data. As with all analog and time-base systems, drift will occur mainly due to component tolerances and temperature dependencies. This inherent system drift limits the absolute accuracy of the measurements that can be attained. The reference data are used primarily as a metric for overall accuracy. Abnormal drift, as would be caused by battery depletion or component degradation, would trigger a system check and data review. The hardware problem would be corrected and field data acquisition would resume. Any previous data deemed degraded would be reacquired. We also plan to preprocess data overnight or concurrent with data acquisition to visually ensure that there are no serious “glitches” or “tears” in the data. Any corrupted lines will be repeated. The GPS will be referenced to a local geodetic marker.

Overview of Quality Assurance (QA): As mentioned above, the planned along-line data density will be around 15 to 20 cm with a line spacing of 50 cm. This will ensure that the 1-m square antennas will sample over every point on the ground. The basic position accuracy of our real-time kinematic differential GPS system is better than 2 cm when operating in “fixed” mode. The LabVIEW program reads the GPS data and mode. If the mode is not fixed, the LabVIEW program flashes a visual warning on the monitor to alert the operator that the GPS is not in fixed mode. Other sources of error in positioning, such as GPS data latency, GPS antenna-to-sensor offset, and tilting of the GPS antenna mast with topography degrade absolute position accuracy. We have added an Attitude Heading and Reference System (AHRS) to measure the cart orientation relative to the ground. We have also developed a navigation program in LabView that runs concurrent with the acquisition program to maintain position over large distances.

Data will also be processed in the field. At the end of each survey line, the data is automatically copied to an external hard drive which will be swapped out with another drive periodically during the survey. The data is then quickly batch processed in Geosoft Oasis Montaj and within minutes the quality of the survey data density and areal coverage can be evaluated.

2.1.6 Additional Records

The following record(s) by this vendor can be accessed via the Internet as Microsoft Word documents at www.uxotestsites.org.

2.2 YPG SITE INFORMATION

2.2.1 Location

YPG is located adjacent to the Colorado River in the Sonoran Desert. The UXO Standardized Test Site is located south of Pole Line Road and east of the Countermine Testing and Training Range. The open field range, calibration grid, blind grid, mogul area, and desert extreme area comprise the 350- by 500-meter general test site area. The open field site is the largest of the test sites and measures approximately 200 by 350 meters. To the east of the open field range are the calibration and blind test grids that measure 30 by 40 meters and 40 by 40 meters, respectively. South of the open field is the 135- by 80-meter mogul area consisting of a sequence of man-made depressions. The desert extreme area is located southeast of the open field site and has dimensions of 50 by 100 meters. The desert extreme area, covered with desert-type vegetation, is used to test the performance of different sensor platforms in a more severe desert conditions/environment.

2.2.2 Soil Type

Soil samples were collected at the YPG UXO Standardized Test Site by ERDC to characterize the shallow subsurface (< 3 m). Both surface grab samples and continuous soil borings were acquired. The soils were subjected to several laboratory analyses, including sieve/hydrometer, water content, magnetic susceptibility, dielectric permittivity, X-ray diffraction, and visual description.

Two soil complexes are present within the site: Riverbend-Carrizo and Cristobal-Gunsight. The Riverbend-Carrizo complex is composed of mixed stream alluvium, whereas the Cristobal-Gunsight complex is derived from fan alluvium. The Cristobal-Gunsight complex covers the majority of the site. Most of the soil samples were classified as either a sandy loam or loamy sand, with most samples containing gravel-size particles. All samples had a measured water content less than 7 percent, except for two that contained 11-percent moisture. The majority of soil samples had water content between 1 and 2 percent. Samples containing more than 3 percent were generally deeper than 1 meter.

An X-ray diffraction analysis on four soil samples indicated a basic mineralogy of quartz, calcite, mica, feldspar, magnetite, and some clay. The presence of magnetite imparted a moderate magnetic susceptibility, with volume susceptibilities generally greater than 100 by 105 SI.

For more details concerning the soil properties at the YPG test site, go to www.uxotestsites.org on the Web to view the entire soils description report.

2.2.3 Test Areas

A description of the test site areas at YPG is included in Table 2.

TABLE 2. TEST SITE AREAS

Area	Description
Calibration grid	Contains the 15 standard ordnance items buried in six positions at various angles and depths to allow demonstrator equipment calibration.
Blind grid	Contains 400 grid cells in a 0.16-hectare (0.39-acre) site. The center of each grid cell contains ordnance, clutter, or nothing.

SECTION 3. FIELD DATA

3.1 DATE OF FIELD ACTIVITIES (17 through 20, 25, and 27 February 2009)

3.2 AREAS TESTED/NUMBER OF HOURS

Areas tested and total number of hours operated at each site are summarized in Table 3.

TABLE 3. AREAS TESTED AND NUMBER OF HOURS

Area	Number of Hours
Calibration lanes	8.50
Blind grid	17.83

3.3 TEST CONDITIONS

3.3.1 Weather Conditions

A YPG weather station located approximately 1 mile west of the test site was used to record average temperature and precipitation on a half-hour basis for each day of operation. The temperatures listed in Table 4 represent the average temperature during field operations from 0700 to 1700 hours, while precipitation data represent a daily total amount of rainfall. Hourly weather logs used to generate this summary are provided in Appendix B.

TABLE 4. TEMPERATURE/PRECIPITATION DATA SUMMARY

Date, 2009	Average Temperature, ^{o}F	Total Daily Precipitation, in.
17 February	58.7	0.00
18 February	58.9	0.00
19 February	64.7	0.00
20 February	66.0	0.00
25 February	74.5	0.00
27 February	70.2	0.00

3.3.2 Field Conditions

USGS surveyed the blind grid on 19, 20, and 25 February 2009. The weather was seasonable, and the field was dry during the survey.

3.3.3 Soil Moisture

Three soil probes were placed at various locations within the site to capture soil moisture data: calibration, mogul, open field, and desert extreme areas. Measurements were collected in percent moisture and were taken twice daily (morning and afternoon) from five different soil depths (1 to 6 in., 6 to 12 in., 12 to 24 in., 24 to 36 in., and 36 to 48 in.) from each probe. Soil moisture logs are included in Appendix C.

3.4 FIELD ACTIVITIES

3.4.1 Setup/Mobilization

These activities included initial mobilization and daily equipment preparation and breakdown. A three-person crew took 9 hours to perform the initial setup and mobilization. There was 1 hour and 20 minutes of daily equipment preparation and 55 minutes of end of day equipment breakdown.

3.4.2 Calibration

USGS spent a total of 8 hours and 30 minutes in the calibration lanes, of which 3 hours and 15 minutes were spent collecting data.

3.4.3 Downtime Occasions

Occasions of downtime are grouped into five categories: equipment/data checks or equipment maintenance, equipment failure and repair, weather, demonstration site issues, or breaks/lunch. All downtime is included for the purposes of calculating labor costs (section 5) except for downtime due to demonstration site issues. Demonstration site issues, while noted in the daily log, are considered non-chargeable downtime for the purposes of calculating labor costs and are not discussed. Breaks and lunches are discussed in this section and billed to the total site survey area.

3.4.3.1 Equipment/data checks, maintenance. Equipment data checks and maintenance activities accounted for no site usage time. These activities included changing out batteries and performing routine data checks to ensure the data were being properly recorded/collected. USGS spent an additional 2 hours and 35 minutes for breaks and lunches.

3.4.3.2 Equipment failure or repair. No time was needed to resolve equipment failures that occurred while surveying the blind grid.

3.4.3.3 Weather. No weather delays occurred during the survey.

3.4.4 Data Collection

USGS spent a total time of 17 hours and 50 minutes in the blind grid area, of which 13 hours were spent collecting data.

3.4.5 Demobilization

The USGS survey crew went on to conduct a full demonstration of the site. Therefore, demobilization did not occur until 27 February 2009. On that day, it took the crew 4 hours to break down and pack up their equipment.

3.5 PROCESSING TIME

USGS submitted the raw data from the demonstration activities on the last day of the demonstration, as required. The scoring submittal data were provided March 2010.

3.6 DEMONSTRATOR'S FIELD PERSONNEL

Ted Asch
Jonah Sullivan
Craig Moulton

3.7 DEMONSTRATOR'S FIELD SURVEYING METHOD

USGS surveyed the blind grid in a linear fashion, in a north-to-south and east-to-west direction.

3.8 SUMMARY OF DAILY LOGS

Daily logs captured all field activities during this demonstration and are located in Appendix D. Activities pertinent to this specific demonstration are indicated in highlighted text.

SECTION 4. TECHNICAL PERFORMANCE RESULTS

4.1 ROC CURVES USING ALL ORDNANCE CATEGORIES

The probability of detection for the response stage (P_d^{res}) and the discrimination stage (P_d^{disc}) versus their respective probability of false positive is shown in Figure 3. Both probabilities plotted against their respective probability of background alarm are shown in Figure 3. Both figures use horizontal lines to illustrate the performance of the demonstrator at two demonstrator-specified points: at the system noise level for the response stage, representing the point below which targets are not considered detectable, and at the demonstrator's recommended threshold level for the discrimination stage, defining the subset of targets the demonstrator would recommend digging based on discrimination. Note that all points have been rounded to protect the ground truth.

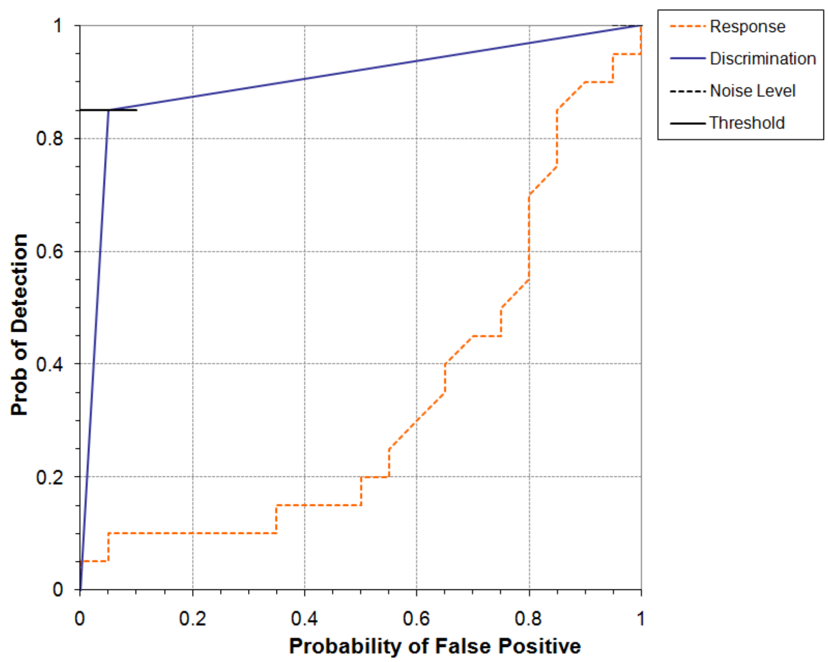

Figure 2. ALLTEM/towed array blind grid probability of detection for response and discrimination stages versus their respective probability of false positive over all ordnance categories combined.

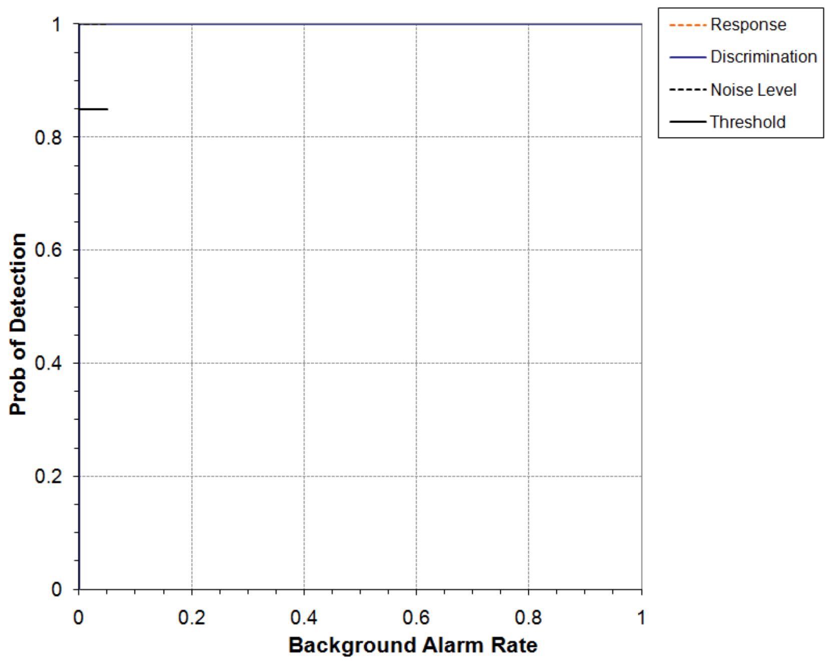

Figure 3. ALLTEM/towed array blind grid probability of detection for response and discrimination stages versus their respective probability of background alarm over all ordnance categories combined.

4.2 ROC CURVES USING ORDNANCE LARGER THAN 20 MM

The probability of detection for the response stage (P_d^{res}) and the discrimination stage (P_d^{disc}) versus their respective probability of false positive when only targets larger than 20 mm are scored is shown in Figure 4. Both probabilities plotted against their respective probability of background alarm are shown in Figure 5. Both figures use horizontal lines to illustrate the performance of the demonstrator at two demonstrator-specified points: at the system noise level for the response stage, representing the point below which targets are not considered detectable, and at the demonstrator's recommended threshold level for the discrimination stage, defining the subset of targets the demonstrator would recommend digging based on discrimination. Note that all points have been rounded to protect the ground truth.

NA

Figure 4. ALLTEM/towed array blind grid probability of detection for response and discrimination stages versus their respective probability of false positive for all ordnance larger than 20 mm.

NA

Figure 5. ALLTEM/towed array blind grid probability of detection for response and discrimination stages versus their respective probabilities of background alarm for all ordnance larger than 20 mm.

4.3 PERFORMANCE SUMMARIES

Results for the blind grid test broken out by size, depth, and nonstandard ordnance are presented in Table 5 (for cost results, see section 5). Results by size and depth include both standard and nonstandard ordnance. The results by size show how well the demonstrator did at detecting/discriminating ordnance of a certain caliber range (see app A for size definitions). The results are relative to the number of ordnance items emplaced. Depth is measured from the geometric center of anomalies.

The RESPONSE STAGE results are derived from the list of anomalies above the demonstrator-provided noise level. The results for the DISCRIMINATION STAGE are derived from the demonstrator's recommended threshold for optimizing UXO field cleanup by minimizing false digs and maximizing ordnance recovery. The lower 90-percent confidence limit on probability of detection and P_{fp} was calculated assuming that the number of detections and false positives are binomially distributed random variables. All results in Table 5 have been rounded to protect the ground truth. However, lower and upper confidence limits were calculated using actual results.

TABLE 5. SUMMARY OF BLIND GRID RESULTS FOR THE ALLTEM

Metric	Overall	Standard	Nonstandard	By Size: Small	Medium	Large	By Depth, m: < 0.3	0.3 to <1	≥ 1
RESPONSE STAGE									
P_d	1.00	1.00	0.95	1.00	1.00	0.95	1.00	1.00	0.85
P_d Low 90% Conf	0.95	0.96	0.87	0.95	0.91	0.75	0.95	0.92	0.55
P_d Upper 90% Conf	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99
P_{fp}	1.00	-	-	-	-	-	1.00	1.00	NA
P_{fp} Low 90% Conf	0.95	-	-	-	-	-	0.94	0.93	-
P_{fp} Upper 90% Conf	1.00	-	-	-	-	-	1.00	1.00	-
P_{ba}	0.00	-	-	-	-	-	-	-	-
DISCRIMINATION STAGE									
P_d	0.85	0.85	0.85	0.90	0.85	0.80	0.90	0.80	0.55
P_d Low 90% Conf	0.78	0.78	0.70	0.79	0.68	0.58	0.83	0.69	0.28
P_d Upper 90% Conf	0.90	0.92	0.91	0.94	0.92	0.92	0.96	0.91	0.83
P_{fp}	0.05	-	-	-	-	-	0.05	0.00	NA
P_{fp} Low 90% Conf	0.03	-	-	-	-	-	0.04	0.00	-
P_{fp} Upper 90% Conf	0.09	-	-	-	-	-	0.12	0.07	-
P_{ba}	0.00	-	-	-	-	-	-	-	-

Response Stage Noise Level: 2.5
Recommended Discrimination Stage Threshold: 0.3

Note: The recommended discrimination stage threshold values are provided by the demonstrator.

4.4 EFFICIENCY, REJECTION RATES, AND TYPE CLASSIFICATION

Efficiency and rejection rates are calculated to quantify the discrimination ability at specific points of interest on the ROC curve: (1) at the point where no decrease in P_d is suffered (i.e., the efficiency is by definition equal to one) and (2) at the operator-selected threshold. These values are reported in Table 6.

TABLE 6. EFFICIENCY AND REJECTION RATES

	Efficiency (E)	False Positive Rejection Rate	Background Alarm Rejection Rate
At operating point	0.86	0.95	NA
With no loss of P_d	1.00	0.00	NA

At the demonstrator's recommended setting, the ordnance items that were detected and correctly discriminated were further scored on whether their correct type could be identified (table 7). Correct type examples include "20-mm projectile, 105-mm HEAT projectile, and 2.75 in. Rocket." A list of the standard type declaration required for each ordnance item was provided to demonstrators prior to testing. For example, the standard types for the three example items are 20 mm, 105 H, and 2.75 in., respectively.

TABLE 7. CORRECT TYPE CLASSIFICATION OF TARGETS CORRECTLY DISCRIMINATED AS UXO

Size	Percentage Correct
Small	0.86
Medium	0.78
Large	0.79
Overall	0.82

4.5 LOCATION ACCURACY

The mean location error and standard deviations are presented in Table 8. These calculations are based on average missed depth for ordnance correctly identified in the discrimination stage. Depths are measured from the closest point of the ordnance to the surface. For the blind grid, only depth errors are calculated, since (X, Y) positions are known to be the centers of each grid square.

TABLE 8. MEAN LOCATION ERROR AND STANDARD DEVIATION

	Mean	Standard Deviation
Depth, m	0.00	0.07

SECTION 5. ON-SITE LABOR COSTS

A standardized estimate for labor costs associated with this effort was calculated as follows: the first person at the test site was designated supervisor, the second person was designated data analyst, and the third and following personnel were considered field support. Standardized hourly labor rates were charged by title: supervisor at $95.00/hour, data analyst at $57.00/hour, and field support at $28.50/hour.

Government representatives monitored on-site activity. All on-site activities were grouped into one of ten categories: initial setup/mobilization, daily setup/stop, calibration, data collection, downtime due to break/lunch, downtime due to equipment failure, downtime due to equipment/data checks or maintenance, downtime due to weather, downtime due to demonstration site issue, or demobilization. See Appendix D for the daily activity log. See section 3.4 for a summary of field activities.

The standardized cost estimate associated with the labor needed to perform the field activities is presented in Table 9. Note that calibration time includes time spent in the calibration lanes as well as field calibrations. Site survey time includes daily setup/stop time, collecting data, breaks/lunch, downtime due to equipment/data checks or maintenance, downtime due to failure, and downtime due to weather.

TABLE 9. ON-SITE LABOR COSTS

	No. People	Hourly Wage	Hours	Cost
Initial setup				
Supervisor	1	$95.00	9.00	$855.00
Data analyst	1	57.00	9.00	513.00
Field support	1	28.50	9.00	256.50
Subtotal				**$1624.50**
Calibration				
Supervisor	1	$95.00	8.50	$807.50
Data analyst	1	57.00	8.50	484.50
Field support	1	28.50	8.50	242.25
Subtotal				**$1534.25**
Site survey				
Supervisor	1	$95.00	17.83	$1693.85
Data analyst	1	57.00	17.83	1016.31
Field support	1	28.50	17.83	508.15
Subtotal				**$3218.31**

See notes at end of table.

TABLE 9 (CONT'D)

	No. People	Hourly Wage	Hours	Cost
Demobilization				
Supervisor	1	$95.00	4.00	$380.00
Data analyst	1	57.00	4.00	228.00
Field support	1	28.50	4.00	114.00
Subtotal				**$722.00**
Total				**$7099.06**

Notes: Calibration time includes time spent in the calibration lanes as well as calibration before each data run.

Site survey time includes daily setup/stop time, collecting data, breaks/lunch, and downtime due to system maintenance, failure, and weather.

SECTION 6.   COMPARISON OF RESULTS TO DATE

No comparisons to date.

SECTION 7. APPENDIXES

APPENDIX A. TERMS AND DEFINITIONS

GENERAL DEFINITIONS

Anomaly: Location of a system response deemed to warrant further investigation by the demonstrator for consideration as an emplaced ordnance item.

Detection: An anomaly location that is within R_{halo} of an emplaced ordnance item.

Emplaced Ordnance: An ordnance item buried by the government at a specified location in the test site.

Emplaced Clutter: A clutter item (i.e., non-ordnance item) buried by the government at a specified location in the test site.

R_{halo}: A pre-determined radius about the periphery of an emplaced item (clutter or ordnance) within which a location identified by the demonstrator as being of interest is considered to be a response from that item. If multiple declarations lie within R_{halo} of any item (clutter or ordnance), the declaration with the highest signal output within the R_{halo} will be utilized. For the purpose of this program, a circular halo 0.5 meters in radius will be placed around the center of the object for all clutter and ordnance items less than 0.6 meters in length. When ordnance items are longer than 0.6 meters, the halo becomes an ellipse where the minor axis remains 1 meter and the major axis is equal to the length of the ordnance plus 1 meter.

Small Ordnance: Caliber of ordnance less than or equal to 40 mm (includes 20-mm projectile, 40-mm projectile, submunitions BLU-26, BLU-63, and M42).

Medium Ordnance: Caliber of ordnance greater than 40 mm and less than or equal to 81 mm (includes 57-mm projectile, 60-mm mortar, 2.75 in. Rocket, MK118 Rockeye, 81-mm mortar).

Large Ordnance: Caliber of ordnance greater than 81 mm (includes 105-mm HEAT, 105-mm projectile, 155-mm projectile, 500-pound bomb).

Shallow: Items buried less than 0.3 meter below ground surface.

Medium: Items buried greater than or equal to 0.3 meter and less than 1 meter below ground surface.

Deep: Items buried greater than or equal to 1 meter below ground surface.

Response Stage Noise Level: The level that represents the point below which anomalies are not considered detectable. Demonstrators are required to provide the recommended noise level for the blind grid test area.

Discrimination Stage Threshold: The demonstrator selected threshold level that they believe provides optimum performance of the system by retaining all detectable ordnance and rejecting the maximum amount of clutter. This level defines the subset of anomalies the demonstrator would recommend digging based on discrimination.

Binomially Distributed Random Variable: A random variable of the type which has only two possible outcomes, say success and failure, is repeated for n independent trials with the probability p of success and the probability 1-p of failure being the same for each trial. The number of successes x observed in the n trials is an estimate of p and is considered to be a binomially distributed random variable.

RESPONSE AND DISCRIMINATION STAGE DATA

The scoring of the demonstrator's performance is conducted in two stages. These two stages are termed the RESPONSE STAGE and DISCRIMINATION STAGE. For both stages, the probability of detection (P_d) and the false alarms are reported as receiver operating characteristic (ROC) curves. False alarms are divided into those anomalies that correspond to emplaced clutter items, measuring the probability of false positive (P_{fp}) and those that do not correspond to any known item, termed background alarms.

The RESPONSE STAGE scoring evaluates the ability of the system to detect emplaced targets without regard to ability to discriminate ordnance from other anomalies. For the RESPONSE STAGE, the demonstrator provides the scoring committee with the location and signal strength of all anomalies that the demonstrator has deemed sufficient to warrant further investigation and/or processing as potential emplaced ordnance items. This list is generated with minimal processing (e.g., this list will include all signals above the system noise threshold). As such, it represents the most inclusive list of anomalies.

The DISCRIMINATION STAGE evaluates the demonstrator's ability to correctly identify ordnance as such, and to reject clutter. For the same locations as in the RESPONSE STAGE anomaly list, the DISCRIMINATION STAGE list contains the output of the algorithms applied in the discrimination-stage processing. This list is prioritized based on the demonstrator's determination that an anomaly location is likely to contain ordnance. Thus, higher output values are indicative of higher confidence that an ordnance item is present at the specified location. For electronic signal processing, priority ranking is based on algorithm output. For other systems, priority ranking is based on human judgment. The demonstrator also selects the threshold that the demonstrator believes will provide "optimum" system performance, (i.e., that retains all the detected ordnance and rejects the maximum amount of clutter).

Note: The two lists provided by the demonstrator contain identical numbers of potential target locations. They differ only in the priority ranking of the declarations.

RESPONSE STAGE DEFINITIONS

Response Stage Probability of Detection (P_d^{res}): P_d^{res} = (No. of response-stage detections)/(No. of emplaced ordnance in the test site).

Response Stage False Positive (fp^{res}): An anomaly location that is within R_{halo} of an emplaced clutter item.

Response Stage Probability of False Positive (P_{fp}^{res}): P_{fp}^{res} = (No. of response-stage false positives)/(No. of emplaced clutter items).

Response Stage Background Alarm (ba^{res}): An anomaly in a blind grid cell that contains neither emplaced ordnance nor an emplaced clutter item. An anomaly location in the open field or scenarios that is outside R_{halo} of any emplaced ordnance or emplaced clutter item.

Response Stage Probability of Background Alarm (P_{ba}^{res}): Blind Grid only: P_{ba}^{res} = (No. of response-stage background alarms)/(No. of empty grid locations).

Response Stage Background Alarm Rate (BAR^{res}): Open Field only: BAR^{res} = (No. of response-stage background alarms)/(arbitrary constant).

Note that the quantities P_d^{res}, P_{fp}^{res}, P_{ba}^{res}, and BAR^{res} are functions of t^{res}, the threshold applied to the response-stage signal strength. These quantities can therefore be written as $P_d^{res}(t^{res})$, $P_{fp}^{res}(t^{res})$, $P_{ba}^{res}(t^{res})$, and $BAR^{res}(t^{res})$.

DISCRIMINATION STAGE DEFINITIONS

Discrimination: The application of a signal processing algorithm or human judgment to response-stage data that discriminates ordnance from clutter. Discrimination should identify anomalies that the demonstrator has high confidence correspond to ordnance, as well as those that the demonstrator has high confidence correspond to non-ordnance or background returns. The former should be ranked with highest priority and the latter with lowest.

Discrimination Stage Probability of Detection (P_d^{disc}): P_d^{disc} = (No. of discrimination-stage detections)/(No. of emplaced ordnance in the test site).

Discrimination Stage False Positive (fp^{disc}): An anomaly location that is within R_{halo} of an emplaced clutter item.

Discrimination Stage Probability of False Positive (P_{fp}^{disc}): P_{fp}^{disc} = (No. of discrimination stage false positives)/(No. of emplaced clutter items).

Discrimination Stage Background Alarm (ba^{disc}): An anomaly in a blind grid cell that contains neither emplaced ordnance nor an emplaced clutter item. An anomaly location in the open field or scenarios that is outside R_{halo} of any emplaced ordnance or emplaced clutter item.

Discrimination Stage Probability of Background Alarm (P_{ba}^{disc}): P_{ba}^{disc} = (No. of discrimination-stage background alarms)/(No. of empty grid locations).

Discrimination Stage Background Alarm Rate (BAR^{disc}): BAR^{disc} = (No. of discrimination-stage background alarms)/(arbitrary constant).

Note that the quantities P_d^{disc}, P_{fp}^{disc}, P_{ba}^{disc}, and BAR^{disc} are functions of t^{disc}, the threshold applied to the discrimination-stage signal strength. These quantities can therefore be written as $P_d^{disc}(t^{disc})$, $P_{fp}^{disc}(t^{disc})$, $P_{ba}^{disc}(t^{disc})$, and $BAR^{disc}(t^{disc})$.

RECEIVER-OPERATING CHARACERISTIC (ROC) CURVES

ROC curves at both the response and discrimination stages can be constructed based on the above definitions. The ROC curves plot the relationship between P_d versus P_{fp} and P_d versus BAR or P_{ba} as the threshold applied to the signal strength is varied from its minimum (t_{min}) to its maximum (t_{max}) value.[1] Figure A-1 shows how P_d versus P_{fp} and P_d versus BAR are combined into ROC curves. Note that the "res" and "disc" superscripts have been suppressed from all the variables for clarity.

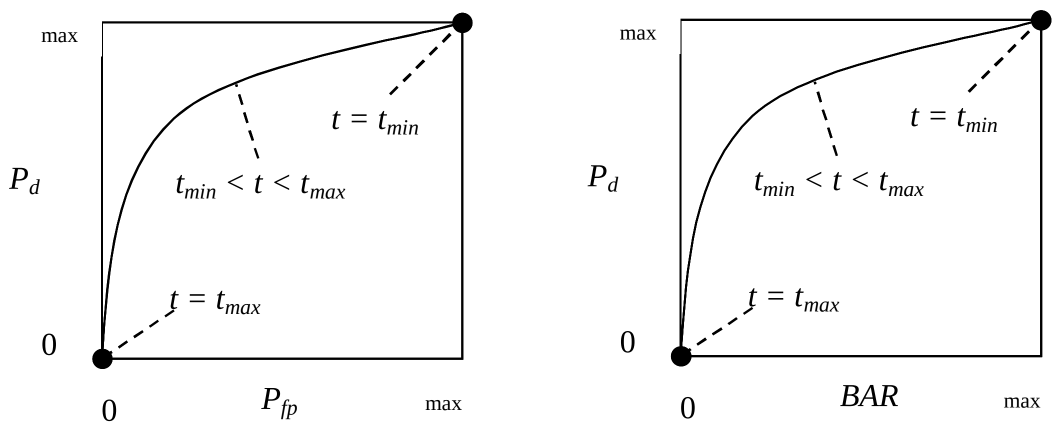

Figure A-1. ROC curves for open field testing. Each curve applies to both the response and discrimination stages.

[1]Strictly speaking, ROC curves plot the P_d versus P_{ba} over a pre-determined and fixed number of detection opportunities (some of the opportunities are located over ordnance and others are located over clutter or blank spots). In an open field scenario, each system suppresses its signal strength reports until some bare-minimum signal response is received by the system. Consequently, the open field ROC curves do not have information from low signal-output locations, and, furthermore, different contractors report their signals over a different set of locations on the ground. These ROC curves are thus not true to the strict definition of ROC curves as defined in textbooks on detection theory. Note, however, that the ROC curves obtained in the blind grid test sites are true ROC curves.

METRICS TO CHARACTERIZE THE DISCRIMINATION STAGE

The demonstrator is also scored on efficiency and rejection ratio, which measure the effectiveness of the discrimination stage processing. The goal of discrimination is to retain the greatest number of ordnance detections from the anomaly list, while rejecting the maximum number of anomalies arising from non-ordnance items. The efficiency measures the amount of detected ordnance retained by the discrimination, while the rejection ratio measures the fraction of false alarms rejected. Both measures are defined relative to the entire response list, i.e., the maximum ordnance detectable by the sensor and its accompanying false positive rate or background alarm rate.

Efficiency (E): $E = P_d^{disc}(t^{disc})/P_d^{res}(t_{min}^{res})$; Measures (at a threshold of interest), the degree to which the maximum theoretical detection performance of the sensor system (as determined by the response stage tmin) is preserved after application of discrimination techniques. Efficiency is a number between 0 and 1. An efficiency of 1 implies that all of the ordnance initially detected in the response stage was retained at the specified threshold in the discrimination stage, t^{disc}.

False Positive Rejection Rate (R_{fp}): $R_{fp} = 1 - [P_{fp}^{disc}(t^{disc})/P_{fp}^{res}(t_{min}^{res})]$; Measures (at a threshold of interest), the degree to which the sensor system's false positive performance is improved over the maximum false positive performance (as determined by the response stage tmin). The rejection rate is a number between 0 and 1. A rejection rate of 1 implies that all emplaced clutter initially detected in the response stage were correctly rejected at the specified threshold in the discrimination stage.

Background Alarm Rejection Rate (R_{ba}):

Blind grid: $R_{ba} = 1 - [P_{ba}^{disc}(t^{disc})/P_{ba}^{res}(t_{min}^{res})]$.
Open field: $R_{ba} = 1 - [BAR^{disc}(t^{disc})/BAR^{res}(t_{min}^{res})])$.

Measures the degree to which the discrimination stage correctly rejects background alarms initially detected in the response stage. The rejection rate is a number between 0 and 1. A rejection rate of 1 implies that all background alarms initially detected in the response stage were rejected at the specified threshold in the discrimination stage.

CHI-SQUARE COMPARISON EXPLANATION:

The Chi-square test for differences in probabilities (or 2 x 2 contingency table) is used to analyze two samples drawn from two different populations to see if both populations have the same or different proportions of elements in a certain category. More specifically, two random samples are drawn, one from each population, to test the null hypothesis that the probability of event A (some specified event) is the same for both populations (ref 3).

A 2 x 2 contingency table is used in the Standardized UXO Technology Demonstration Site Program to determine if there is reason to believe that the proportion of ordnance correctly detected/discriminated by demonstrator X’s system is significantly degraded by the more challenging terrain feature introduced. The test statistic of the 2 x 2 contingency table is the Chi-square distribution with one degree of freedom. Since an association between the more challenging terrain feature and relatively degraded performance is sought, a one-sided test is performed. A significance level of 0.05 is chosen which sets a critical decision limit of 2.71 from the Chi-square distribution with one degree of freedom. It is a critical decision limit because if the test statistic calculated from the data exceeds this value, the two proportions tested will be considered significantly different. If the test statistic calculated from the data is less than this value, the two proportions tested will be considered not significantly different.

An exception must be applied when either a 0 or 100 percent success rate occurs in the sample data. The Chi-square test cannot be used in these instances. Instead, Fischer's test is used and the critical decision limit for one-sided tests is the chosen significance level, which in this case is 0.05. With Fischer's test, if the test statistic is less than the critical value, the proportions are considered to be significantly different.

Standardized UXO Technology Demonstration Site examples, where blind grid results are compared to those from the open field and open field results are compared to those from one of the scenarios, follow. It should be noted that a significant result does not prove a cause and effect relationship exists between the two populations of interest; however, it does serve as a tool to indicate that one data set has experienced a degradation in system performance at a large enough level than can be accounted for merely by chance or random variation. Note also that a result that is not significant indicates that there is not enough evidence to declare that anything more than chance or random variation within the same population is at work between the two data sets being compared.

Demonstrator X achieves the following overall results after surveying each of the three progressively more difficult areas using the same system (results indicate the number of ordnance detected divided by the number of ordnance emplaced):

	Blind grid	Open field	Moguls
P_d^{res}	100/100 = 1.0	8/10 = .80	20/33 = .61
P_d^{disc}	80/100 = 0.80	6/10 = .60	8/33 = .24

P_d^{res}: BLIND GRID versus OPEN FIELD. Using the example data above to compare probabilities of detection in the response stage, all 100 ordnance out of 100 emplaced ordnance items were detected in the blind grid while 8 ordnance out of 10 emplaced were detected in the open field. Fischer's test must be used since a 100 percent success rate occurs in the data. Fischer's test uses the four input values to calculate a test statistic of 0.0075 that is compared against the critical value of 0.05. Since the test statistic is less than the critical value, the smaller response stage detection rate (0.80) is considered to be significantly less at the 0.05 level of significance. While a significant result does not prove a cause and effect relationship exists between the change in survey area and degradation in performance, it does indicate that the detection ability of demonstrator X's system seems to have been degraded in the open field relative to results from the blind grid using the same system.

P_d^{disc}: BLIND GRID versus OPEN FIELD. Using the example data above to compare probabilities of detection in the discrimination stage, 80 out of 100 emplaced ordnance items were correctly discriminated as ordnance in blind grid testing while 6 ordnance out of 10 emplaced were correctly discriminated as such in open field-testing. Those four values are used to calculate a test statistic of 1.12. Since the test statistic is less than the critical value of 2.71, the two discrimination stage detection rates are considered to be not significantly different at the 0.05 level of significance.

P_d^{res}: OPEN FIELD versus MOGULS. Using the example data above to compare probabilities of detection in the response stage, 8 out of 10 and 20 out of 33 are used to calculate a test statistic of 0.56. Since the test statistic is less than the critical value of 2.71, the two response stage detection rates are considered to be not significantly different at the 0.05 level of significance.

P_d^{disc}: OPEN FIELD versus MOGULS. Using the example data above to compare probabilities of detection in the discrimination stage, 6 out of 10 and 8 out of 33 are used to calculate a test statistic of 2.98. Since the test statistic is greater than the critical value of 2.71, the smaller discrimination stage detection rate is considered to be significantly less at the 0.05 level of significance. While a significant result does not prove a cause and effect relationship exists between the change in survey area and degradation in performance, it does indicate that the ability of demonstrator X to correctly discriminate seems to have been degraded by the mogul terrain relative to results from the flat open field using the same system.

APPENDIX B. DAILY WEATHER LOGS

Date, 2009	Time, EST	Average Temperature, °F	Average Precipitation, in.
17 February	0700	47.5	0.00
	0800	48.0	0.00
	0900	53.4	0.00
	1000	56.6	0.00
	1100	58.8	0.00
	1200	61.0	0.00
	1300	63.4	0.00
	1400	64.4	0.00
	1500	64.6	0.00
	1600	64.4	0.00
	1700	64.1	0.00
18 February	0700	41.2	0.00
	0800	41.6	0.00
	0900	48.6	0.00
	1000	56.6	0.00
	1100	58.2	0.00
	1200	62.5	0.00
	1300	65.7	0.00
	1400	66.8	0.00
	1500	69.7	0.00
	1600	69.5	0.00
	1700	68.4	0.00
19 February	0700	48.7	0.00
	0800	50.3	0.00
	0900	54.8	0.00
	1000	60.5	0.00
	1100	64.8	0.00
	1200	67.6	0.00
	1300	70.8	0.00
	1400	72.9	0.00
	1500	74.1	0.00
	1600	74.0	0.00
	1700	73.5	0.00

Date, 2009	Time, EST	Average Temperature, °F	Average Precipitation, in.
20 February	0700	49.5	0.00
	0800	50.3	0.00
	0900	56.9	0.00
	1000	59.8	0.00
	1100	64.0	0.00
	1200	69.2	0.00
	1300	74.0	0.00
	1400	75.3	0.00
	1500	76.0	0.00
	1600	76.2	0.00
	1700	75.3	0.00
25 February	0700	54.1	0.00
	0800	55.8	0.00
	0900	64.4	0.00
	1000	74.6	0.00
	1100	77.8	0.00
	1200	77.7	0.00
	1300	80.5	0.00
	1400	82.5	0.00
	1500	83.7	0.00
	1600	84.3	0.00
	1700	83.7	0.00
27 February	0700	50.7	0.00
	0800	51.7	0.00
	0900	60.2	0.00
	1000	66.3	0.00
	1100	73.4	0.00
	1200	75.4	0.00
	1300	77.2	0.00
	1400	77.9	0.00
	1500	79.5	0.00
	1600	79.9	0.00
	1700	79.5	0.00

APPENDIX C. SOIL MOISTURE

Date: 18 February 2009			
Times: 0700 and 1400			
Probe Location	**Layer, in.**	**AM Reading, %**	**PM Reading, %**
Calibration area	0 to 6	4.2	3.4
	6 to 12	12.9	12.0
	12 to 24	9.1	9.0
	24 to 36	3.6	3.4
	36 to 48	7.5	7.6
Mogul field	0 to 6	0.5	1.4
	6 to 12	2.0	5.7
	12 to 24	9.6	4.2
	24 to 36	10.1	5.7
	36 to 48	7.6	4.5
Desert Extreme area	0 to 6	0.5	1.4
	6 to 12	5.7	38.2
	12 to 24	4.0	8.1
	24 to 36	5.6	4.7
	36 to 48	4.7	4.9

Date: 19 February 2009			
Times: 0700 and 1400			
Probe Location	**Layer, in.**	**AM Reading, %**	**PM Reading, %**
Calibration area	0 to 6	2.8	3.4
	6 to 12	11.7	12.0
	12 to 24	9.3	9.3
	24 to 36	3.8	3.5
	36 to 48	7.5	7.6
Mogul field	0 to 6	2.5	1.4
	6 to 12	6.3	6.0
	12 to 24	6.1	7.1
	24 to 36	7.5	7.6
	36 to 48	10.5	10.3
Desert Extreme area	0 to 6	3.7	3.5
	6 to 12	38.2	38.2
	12 to 24	3.9	3.7
	24 to 36	5.6	5.8
	36 to 48	5.0	4.7

Date: 20 February 2009			
Times: 0700 and 1400			
Probe Location	**Layer, in.**	**AM Reading, %**	**PM Reading, %**
Calibration area	0 to 6	3.7	3.1
	6 to 12	10.9	10.9
	12 to 24	8.8	9.3
	24 to 36	4.0	3.2
	36 to 48	7.6	7.9
Mogul field	0 to 6	2.2	0.8
	6 to 12	38.2	6.3
	12 to 24	8.1	6.5
	24 to 36	4.7	7.6
	36 to 48	4.9	9.3
Desert Extreme area	0 to 6	11.1	11.1
	6 to 12	38.2	38.2
	12 to 24	2.8	3.3
	24 to 36	5.8	6.0
	36 to 48	4.5	4.2

Date: 25 February 2009			
Times: 0900 and 1400			
Probe Location	**Layer, in.**	**AM Reading, %**	**PM Reading, %**
Calibration area	0 to 6	4.0	4.5
	6 to 12	10.3	9.7
	12 to 24	9.1	9.4
	24 to 36	3.4	3.5
	36 to 48	7.6	7.5
Mogul field	0 to 6	2.5	0.0
	6 to 12	2.8	0.8
	12 to 24	5.2	6.9
	24 to 36	6.6	7.1
	36 to 48	9.3	9.9
Desert Extreme area	0 to 6	40.0	11.1
	6 to 12	38.2	38.2
	12 to 24	3.1	8.1
	24 to 36	6.2	6.2
	36 to 48	4.9	4.2

Date: 27 February 2009			
Times: 0800 and NA			
Probe Location	**Layer, in.**	**AM Reading, %**	**PM Reading, %**
Calibration area	0 to 6	3.7	-
	6 to 12	8.9	-
	12 to 24	9.1	-
	24 to 36	3.2	-
	36 to 48	7.6	-
Mogul field	0 to 6	2.5	-
	6 to 12	3.4	-
	12 to 24	6.9	-
	24 to 36	8.3	-
	36 to 48	8.9	-
Desert Extreme area	0 to 6	11.1	-
	6 to 12	38.2	-
	12 to 24	3.3	-
	24 to 36	5.8	-
	36 to 48	4.9	-

(Page C-4 Blank)

Date, 09	No. of People	Area-Tested	Status Start Time	Status Stop Time	Duration, min	Operational Status	Operational Status - Comments	Track Method	Pattern	Field Conditions	
17 Feb	3	CALIBRATION LANES	1300	1630	210	INITIAL SET-UP	Initial Mobilization	GPS	Linear	Sunny	Warm
18 Feb	3	CALIBRATION LANES	700	1230	330	INITIAL SET-UP	Initial Mobilization	GPS	Linear	Sunny	Warm
18 Feb	3	CALIBRATION LANES	1230	1300	30	BREAK/LUNCH	Lunch	GPS	Linear	Sunny	Warm
18 Feb	3	CALIBRATION LANES	1300	1415	75	DOWNTIME DUE TO EQUIP MAINT/CHECK	Checking equipment	GPS	Linear	Sunny	Warm
18 Feb	3	CALIBRATION LANES	1415	1615	120	COLLECTING DATA	Collecting data, West - East, South, North	GPS	Linear	Sunny	Warm
18 Feb	3	CALIBRATION LANES	1615	1630	15	DAILY START, STOP	Breakdown, end of day	GPS	Linear	Sunny	Warm
19 Feb	3	CALIBRATION LANES	700	800	60	DAILY START, STOP	Setting up test equipment	GPS	Linear	Sunny	Warm
19 Feb	3	CALIBRATION LANES	800	1015	135	DOWNTIME DUE TO EQUIPMENT FAILURE	Software and GPS issues	GPS	Linear	Sunny	Warm
19 Feb	3	CALIBRATION LANES	1015	1130	75	COLLECTING DATA	Collecting data, West - East, South, North	GPS	Linear	Sunny	Warm
19 Feb	3	BLIND TEST GRID	1130	1230	60	COLLECTING DATA	Collecting data, West - East, South, North	GPS	Linear	Sunny	Warm
19 Feb	3	BLIND TEST GRID	1230	1300	30	BREAK/LUNCH	Lunch	GPS	Linear	Sunny	Warm
19 Feb	3	BLIND TEST GRID	1300	1540	160	COLLECTING DATA	Collecting data, West - East, South, North	GPS	Linear	Sunny	Warm
20 Feb	3	BLIND TEST GRID	1540	1610	30	DAILY START, STOP	Breakdown, end of day	GPS	Linear	Sunny	Warm
20 Feb	3	BLIND TEST GRID	700	820	80	DAILY START, STOP	Setting up test equipment	GPS	Linear	Sunny	Warm
20 Feb	3	BLIND TEST GRID	820	1140	200	COLLECTING DATA	Collecting data, West - East, South, North	GPS	Linear	Sunny	Warm
20 Feb	3	BLIND TEST GRID	1140	1215	35	BREAK/LUNCH	Lunch	GPS	Linear	Sunny	Warm
20 Feb	3	OPEN FIELD	1215	1545	210	COLLECTING DATA	Collecting Data, North - South, West - East	GPS	Linear	Sunny	Warm
20 Feb	3	OPEN FIELD	1545	1615	30	DAILY START, STOP	Breakdown end of day	GPS	Linear	Sunny	Warm
21 Feb	3	OPEN FIELD	700	810	70	DAILY START, STOP	Setting up test equipment	GPS	Linear	Sunny	Warm
21 Feb	3	OPEN FIELD	810	1120	190	COLLECTING DATA	Collecting Data, North - South, West - East	GPS	Linear	Sunny	Warm
21 Feb	3	OPEN FIELD	1120	1150	30	BREAK/LUNCH	Lunch	GPS	Linear	Sunny	Warm
21 Feb	3	OPEN FIELD	1150	1615	265	COLLECTING DATA	Collecting Data, North - South, West - East	GPS	Linear	Sunny	Warm
21 Feb	3	OPEN FIELD	1615	1630	15	DAILY START, STOP	Breakdown, end of day	GPS	Linear	Sunny	Warm

Date, 09	No. of People	Area-Tested	Status Start Time	Status Stop Time	Duration, min.	Operational Status	Operational Status - Comments	Track Method	Pattern	Field Conditions	
23 Feb	3	OPEN FIELD	715	815	60	DAILY START, STOP	Setting up test equipment	GPS	Linear	Sunny	Warm
23 Feb	3	OPEN FIELD	815	1130	195	COLLECTING DATA	Collecting Data, North - South, West - East	GPS	Linear	Sunny	Warm
23 Feb	3	OPEN FIELD	1130	1300	90	COLLECTING DATA	Collecting Data, North - South, West - East	GPS	Linear	Sunny	Warm
23 Feb	3	OPEN FIELD	1130	1300	90	BREAK/LUNCH	Lunch	GPS	Linear	Sunny	Warm
23 Feb	3	OPEN FIELD	1300	1630	210	COLLECTING DATA	Collecting Data, North - South, West - East	GPS	Linear	Sunny	Warm
23 Feb	3	OPEN FIELD	1630	1645	15	DAILY START, STOP	Breakdown, end of day	GPS	Linear	Sunny	Warm
24 Feb	3	OPEN FIELD	700	745	45	DAILY START, STOP	Set up of equipment	GPS	Linear	Sunny	Warm
24 Feb	3	OPEN FIELD	745	1200	255	COLLECTING DATA	Collecting Data, North - South, West - East	GPS	Linear	Sunny	Warm
24 Feb	3	OPEN FIELD	1200	1330	90	COLLECTING DATA	Collecting Data, North - South, West - East	GPS	Linear	Sunny	Warm
24 Feb	3	OPEN FIELD	1200	1330	90	BREAK/LUNCH	Lunch	GPS	Linear	Sunny	Warm
24 Feb	3	OPEN FIELD	1330	1520	110	COLLECTING DATA	Collecting Data, North - South, West - East	GPS	Linear	Sunny	Warm
24 Feb	3	OPEN FIELD	1520	1545	25	COLLECTING DATA	Collecting Data, South - North, West - East	GPS	Linear	Sunny	Warm
24 Feb	3	OPEN FIELD	1545	1600	15	DAILY START, STOP	Breakdown, end of day	GPS	Linear	Sunny	Warm
25 Feb	3	OPEN FIELD	900	945	45	DAILY START, STOP	Set up of equipment	GPS	Linear	Sunny	Warm
25 Feb	3	BLIND TEST GRID	945	1220	155	COLLECTING DATA	Collecting Data, South - North, West - East	GPS	Linear	Sunny	Warm
25 Feb	3	BLIND TEST GRID	1220	1350	90	BREAK/LUNCH	Lunch	GPS	Linear	Sunny	Warm
25 Feb	3	BLIND TEST GRID	1220	1350	90	COLLECTING DATA	Collecting Data, North - South, West - East	GPS	Linear	Sunny	Warm
25 Feb	3	BLIND TEST GRID	1350	1545	115	COLLECTING DATA	Collecting Data, North - South, West - East	GPS	Linear	Sunny	Warm
25 Feb	3	BLIND TEST GRID	1545	1610	25	DAILY START, STOP	Breakdown, end of day	GPS	Linear	Sunny	Warm
26 Feb	3	OPEN FIELD	715	751	36	DAILY START, STOP	Setup of equipment	GPS	Linear	Sunny	Warm
26 Feb	3	OPEN FIELD	751	1600	489	COLLECTING DATA	Collecting Data, North - South, West - East	GPS	Linear	Sunny	Warm
26 Feb	3	OPEN FIELD	1600	1620	20	DAILY START, STOP	Breakdown end of day	GPS	Linear	Sunny	Warm
27 Feb	3	OPEN FIELD	700	745	45	DAILY START, STOP	Setup of equipment	GPS	Linear	Sunny	Warm
27 Feb	3	OPEN FIELD	745	1100	195	COLLECTING DATA	Collecting Data, North - South, West - East	GPS	Linear	Sunny	Warm
27 Feb	3	OPEN FIELD	1100	1500	240	DEMOBILIZATION	Demobilization	GPS	Linear	Sunny	Warm

Note: Activities pertinent to this specific demonstration are indicated in highlighted text.

APPENDIX E. REFERENCES

1. Standardized UXO Technology Demonstration Site Handbook, DTC Project No. 8-CO-160-000-473, Report No. ATC-8349, March 2002.

2. Aberdeen Proving Ground Soil Survey Report, October 1998.

3. Data Summary, UXO Standardized Test Site: APG Soils Description, May 2002.

4. Yuma Proving Ground Soil Survey Report, May 2003.

5. Practical Nonparametric Statistics, W.J. Conover, John Wiley & Sons, 1980, pages 144 through 151.

APPENDIX F. ABBREVIATIONS

ADST	=	Aberdeen Data Services Team
APG	=	Aberdeen Proving Ground
ATC	=	U.S. Army Aberdeen Test Center
ATSS	=	Aberdeen Test Support Services
E	=	efficiency
ERDC	=	U.S. Army Corps of Engineers Engineering Research and Development Center
ESTCP	=	Environmental Security Technology Certification Program
EQT	=	Army Environmental Quality Technology Program
GPS	=	Global Positioning System
HDSD	=	Homeland Defense and Sustainment Division
HEAT	=	high-explosive antitank
JPG	=	Jefferson Proving Ground
M	=	standard deviation
NS	=	nonstandard
POC	=	point of contact
QA	=	quality assurance
QC	=	quality control
ROC	=	receiver-operating characteristic
SERDP	=	Strategic Environmental Research and Development Program
SL	=	Survivability and Lethality
USAEC	=	U.S. Army Environmental Command
UXO	=	unexploded ordnance
YPG	=	U.S. Army Yuma Proving Ground

APPENDIX G. DISTRIBUTION LIST

DTC Project No.8-CO-160-UXO-021

Addressee	No. of Copies
Commander U.S. Army Aberdeen Test Center ATTN: TEDT-AT-SLE (Mr. J. Stephen McClung) 400 Colleran Road Aberdeen Proving Ground, MD 21005-5059	1
Program Manager SERDP/ESTCP Munitions Management ATTN: Mr. Herb Nelson 901 North Stuart Street, Suite 303 Arlington, VA 22203	1
U.S. Geological Survey ATTN: Mr. Ted Asch Denver Federal Center Bldg. 20, MS-964 Denver, CO 80225-0046	1
Defense Technical Information Center 8725 John J. Kingman Road, STE 0944 Fort Belvoir, VA 22060-6218	PDF

Secondary distribution is controlled by Program Manager, SERDP/ESTCP, Munitions Management, ATTN: Mr. Herb Nelson.